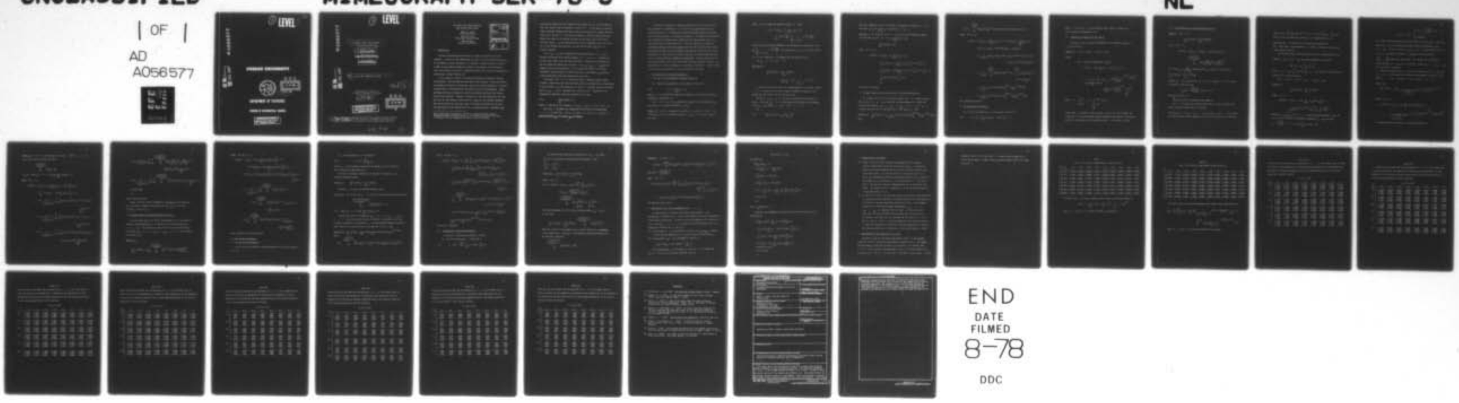
AD-A056 577 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
ON SUBSET SELECTION PROCEDURES FOR POISSON POPULATIONS,(U)
MAY 78 S S GUPTA, Y LEONG, W WONG N00014-75-C-0455
UNCLASSIFIED MIMEOGRAPH-SER-78-3 NL

| OF |

AD
A056577



END
DATE
FILMED
8-78
DDC

(1) LEVEL II

AD A056577

AD No.
DDC FILE COPY

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

6

ON SUBSET SELECTION PROCEDURES FOR POISSON POPULATIONS

10 Shanti S. Gupta,
~~Purdue University~~

Yoon-Kwai Leong
~~Tunku Abdul Rahman College~~

Wing-Yue Wong
[redacted]

14 MIMEOGRAPH-SER-78-3

12 35p.

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-3

11 May 1978

15 *This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

291 730

ON SUBSET SELECTION PROCEDURES
FOR POISSON POPULATIONS*

Shanti S. Gupta
Purdue University

Yoon-Kwai Leong
Tunku Abdul Rahman College

Wing-Yue Wong
University of Malaya

1. Introduction

Poisson distribution has been used as a model in several statistical problems. As early as 1898, Bortkiewicz [1] used it to fit the data pertaining to the deaths by kicks from horses in a regiment. Poisson process is used as a model in many applied probability problems, for example, for the waiting time, for arrivals of calls at a telephone exchange, for arrivals of radioactive particles at a Geiger counter, etc.

In this paper our object is to study the problem of comparing k Poisson distributions. Not much work has been done on this problem. More specifically, we consider the problem of selecting a subset of k Poisson populations including the best which is associated with the smallest value of the paramater. Gupta and Huang [4] have considered the selection problem according to the largest value of the parameter. However, a procedure of the type proposed by them does not work for the problem of selection with respect to the smallest parameter. Goel [3] has shown that the usual type of selection procedures do not exist for some values of the probability P* of a correct selection. Moreover Leong

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

and Wong [6] showed that the infimum of the probability of a correct selection when the location type of procedure is used in k^{-1}. In this paper, we propose some procedures different from that of Gupta and Huang [4] for subset selection which exist for all P*. The rules are based on a result of Chapman [2] who showed that there is no unbiased estimator of the ratio $\lambda_1\lambda_2^{-1}$ with finite variance, where λ_1, λ_2 are expected values of two independent random variables X_1, X_2 with Poissòn distributions, but that the estimator $X_1(X_2+1)^{-1}$ is "almost unbiased".

Let $\pi_1,\pi_2,\ldots,\pi_k$ be k independent Poisson populations, i.e., π_i has a Poisson distribution with unknown parameter λ_i, $i = 1,2,\ldots,k$. Suppose that we have equal sample size from each population. Without loss of generality, one can assume the sample size to be one. Let $\lambda_{[1]} \leq \lambda_{[2]} \leq \cdots \leq \lambda_{[k]}$ be the ordered values of the parameters; it is assumed that there is no a priori information available about the correct pairing of the ordered $\lambda_{[i]}$ and the k given populations from which observations are taken.

Given any P* ($\frac{1}{k} < P^* < 1$), we wish to select a nonempty (small) subset of these k populations such that the subset contains the population corresponding to the parameter $\lambda_{[1]}$ with probability at least P*, no matter what the configuration of $\lambda_1,\lambda_2,\ldots,\lambda_k$ is. We denote this notation by CS. Therefore we are interested in defining a selection procedure R such that

$$\inf_{\underset{\sim}{\lambda}\in\Omega} P_{\underset{\sim}{\lambda}}(CS|R) \geq P^* \tag{1.1}$$

where Ω is the set of all k-tuples $\underset{\sim}{\lambda} = (\lambda_1,\lambda_2,\ldots,\lambda_k)$, $\lambda_i > 0$, $i = 1,2,\ldots,k$.

Let $X_1,X_2,\ldots,X_k$ denote the independent observations from populations $\pi_1,\pi_2,\ldots,\pi_k$, respectively. Let $X_{(i)}$ be that value of $X_1,X_2,\ldots,X_k$ which is associated with $\lambda_{[i]}$; of course $\lambda_{[i]}$ is unknown.

In Section 2, we discuss a subset selection rule so as to satisfy the basic probability requirement (1.1), and to find an upper bound for the expected subset size. A conditional selection procedure based on the total sum of the observations is considered in Section 3. A method for constructing the conservative constants and an upper bound for the expected subset size are derived for this conditional rule. Section 4 deals with a different selection procedure of the type suggested by Seal for the normal mean problem. We also discuss the Seal type procedure conditioning on the total sum of the observation, in which case the selection constant can be determined precisely so as to satisfy the basic probability requirement. An exact expression for the expected subset size of the conditional Seal type procedure is given. An application to a test of homogeneity is mentioned in Section 5. Tables related to the selection procedures are given at the end of the paper.

2. The Unconditional Selection Procedure R_1

2.1. The Rule R_1 and Probability of Correct Selection

R_1: Select the population π_i in the subset if and only if

$$X_i \leq c_1 \min_{1 \leq j \leq k} X_j + c_1 \tag{2.1}$$

where $c_1 \geq 1$ is the smallest number to be chosen so as to satisfy the basic probability requirement (1.1).

For $i = 1,2,\ldots,k$, let $\pi_{(i)}$ denote the population associated with $\lambda_{[i]}$ and let $p_{\underline{\lambda}}(i) = P_{\underline{\lambda}}(\text{select population } \pi_{(i)} | R_1)$.

Theorem 2.1. $p_{\underline{\lambda}}(i)$ is a decreasing function in $\lambda_{[i]}$ when all other λ's are fixed and $p_{\underline{\lambda}}(i)$ is an increasing function in $\lambda_{[j]}$, $j \neq i$, when all other λ's are fixed.

Proof. Let $<x>$ denote the smallest integer $\geq x$. Then

$$p_{\underline{\lambda}}(i) = P_{\underline{\lambda}}(X_{(i)} \leq c_1 \min_{1 \leq j \leq k} X_{(j)} + c_1)$$

$$= \sum_{x=0}^{\infty} e^{-\lambda_{[i]}} \frac{\lambda_{[i]}^x}{x!} \{ \prod_{\substack{j=1 \\ j \neq i}}^{k} \sum_{\ell = < \frac{x}{c_1} - 1 >}^{\infty} e^{-\lambda_{[j]}} \frac{\lambda_{[j]}^{\ell}}{\ell!} \}.$$

Since Poisson distribution belongs to the stochastically increasing family and $\prod_{\substack{j=1 \\ j \neq i}}^{k} \{ \sum_{\ell = < \frac{x}{c_1} - 1 >} e^{-\lambda_{[j]}} \frac{\lambda_{[j]}^{\ell}}{\ell!} \}$ is increasing in $\lambda_{[j]}$, $j \neq i$ and decreasing in x, so by a lemma on p. 112 Lehmann [5], the results follow.

Let $\Omega_0 = \{\underline{\lambda} = (\lambda,\ldots,\lambda): \lambda > 0\}$.

Corollary 2.1.

$$\inf_{\underline{\lambda} \in \Omega} P_{\underline{\lambda}}(CS|R_1) = \inf_{\underline{\lambda} \in \Omega_0} P_{\underline{\lambda}}(CS|R_1)$$

$$= \inf_{\lambda > 0} \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} \{ \sum_{\ell = < \frac{x}{c_1} - 1 >} e^{-\lambda} \frac{\lambda^{\ell}}{\ell!} \}^{k-1}.$$

It should be pointed out that the infimum depends on the common unknown λ, $\lambda > 0$. In Section 6, we discuss numerical methods to determine this infimum and the constant for the selection rule.

For any $\underline{\lambda} = (\lambda_1,\ldots,\lambda_k) \in \Omega$, the joint conditional distribution of $X_1, X_2, \ldots, X_k$ given $\sum_{i=1}^{k} X_i = t$ is a multinomial distribution with parameters t and $(\theta_1,\ldots,\theta_k)$ where $\theta_i = \lambda_i(\lambda_1+\ldots+\lambda_k)^{-1}$, $i = 1,\ldots,k$.

Let

$$A(k,t,c_1(t)) = \sum \frac{t!}{x_1! \ldots x_k!} \left(\frac{1}{k}\right)^t \tag{2.2}$$

where the summation is over all k-tuples of nonnegative integers $(x_1,\ldots,x_k)$ such that $x_1 \le c_1(t) \min_{2\le j\le k} x_j + c_1(t)$ and $x_1+\ldots+x_k = t$.

Theorem 2.2. For given P*, any t, $t \ge 0$, let $c_1(t)$ be the smallest number such that $A(k,t,c_1(t)) \ge P^*$. If $c_1 = \sup_{t\ge 0}\{c_1(t)\}$, then

$$\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_1) \ge P^*.$$

Proof. For $\underline{\lambda} \in \Omega_0$,

$$P_{\underline{\lambda}}(CS|R_1) = P_{\underline{\lambda}}(X_{(1)} \le c_1 \min_{2\le j\le k} X_{(j)} + c_1)$$

$$= \sum_{t=0}^{\infty} P_{\underline{\lambda}}(X_{(1)} \le c_1 \min_{2\le j\le k} X_{(j)}+c_1 | \sum_{i=1}^{k} X_i=t)P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i=t)$$

$$\ge \sum_{t=0}^{\infty} P_{\underline{\lambda}}(X_{(1)} \le c_1(t) \min_{2\le j\le k} X_{(j)}+c_1(t)| \sum_{i=1}^{k} X_i=t)P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i=t)$$

$$= \sum_{t=0}^{\infty} A(k,t,c_1(t))P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i=t)$$

$$\ge P^*.$$

This proves the theorem.

2.2. An Upper Bound on the Expected Subset Size Associated with R_1.

Let S denote the size of the selected subset, then S is a random variable taking value 1,2,...,k. Let us consider the expected values of S under the slippage configuration $\lambda_{[1]} = \delta\lambda$, $\lambda_{[2]} = \ldots = \lambda_{[k]} = \lambda$, $0 < \delta < 1$, $0 < \lambda_0 < \lambda$. We denote the space of all configurations of this type by Ω_1. Then

Theorem 2.3. $$\sup_{\underline{\lambda}\in\Omega_1} E_{\underline{\lambda}}(S|R_1) \le k- \inf_{t\ge[c_1]+1} \{g(t,\delta)+(k-1)g(t,\tfrac{1}{\delta})\}\int_0^{(1+\delta)\lambda_0} \frac{1}{[c_1]!}\, y^{[c_1]} e^{-y}dy$$

where

$$g(t,\delta) = \sum_{i=0}^{[\frac{t-c_1}{1+c_1}]} \binom{t}{i}\left(\frac{1}{1+\delta}\right)^i\left(\frac{\delta}{1+\delta}\right)^{t-i}$$ and $[x]$ denote the integral part of x.

Proof. For $\underline{\lambda} \in \Omega_1$,

$$E_{\underline{\lambda}}(S|R_1) = P_{\underline{\lambda}}(X_{(1)} \leq c_1 \min_{2\leq i\leq k} X_{(i)}+c_1) + (k-1)P_{\underline{\lambda}}(X_{(2)}\leq c_1 \min_{\substack{1\leq i\leq k\\ i\neq 2}} X_{(i)}+c_1)$$

$$\leq P_{\underline{\lambda}}(X_{(1)} \leq c_1X_{(2)}+c_1)+(k-1)P_{\underline{\lambda}}(X_{(2)} \leq c_1X_{(1)}+c_1)$$

$$= k- \sum_{t=[c_1]+1}^{\infty} \{P_{\underline{\lambda}}(X_{(1)}>c_1X_{(2)}+c_1|X_{(1)}+X_{(2)}=t)+(k-1)P_{\underline{\lambda}}(X_{(2)}>c_1X_{(1)}+c_1|X_{(1)}+X_{(2)}=t)\}P_{\underline{\lambda}}(X_{(1)}+X_{(2)}=t)$$

$$= k- \sum_{t=[c_1]+1}^{\infty} \{ \sum_{i=0}^{[\frac{t-c_1}{1+c_1}]} \binom{t}{i}\left(\frac{1}{1+\delta}\right)^i\left(\frac{\delta}{1+\delta}\right)^{t-i}+(k-1) \sum_{i=0}^{[\frac{t-c_1}{1+c_1}]} \binom{t}{i}\left(\frac{\delta}{1+\delta}\right)^i\left(\frac{1}{1+\delta}\right)^{t-i}\} e^{-(1+\delta)\lambda}\frac{((1+\delta)\lambda)^t}{t!}$$

$$\leq k- \inf_{t\geq[c_1]+1} \{g(t,\delta)+(k-1)g(t,\tfrac{1}{\delta})\}\int_0^{(1+\delta)\lambda} \frac{1}{[c_1]!}\, y^{[c_1]} e^{-y}dy$$

$$\leq k- \inf_{t\geq[c_1]+1} \{g(t,\delta)+(k-1)g(t,\tfrac{1}{\delta})\}\int_0^{(1+\delta)\lambda_0} \frac{1}{[c_1]!}\, y^{[c_1]} e^{-y}dy.$$

This completes the proof.

3. The Conditional Procedure R_2

R_2: Select the population π_i in the subset if and only if

(3.1) $$X_i \leq c_2(t) \min_{1\leq j\leq k} X_j+c_2(t), \quad \text{given} \sum_{i=1}^{k} X_i = t$$

where $t \geq 0$ and $c_2(t) \geq 1$ is the smallest value chosen to satisfy the basic probability requirement (1.1).

3.1. Monotonicity property for the rule R_2

As before, let $p_{\underline{\lambda}}(i)$ denote the probability of selecting population $\pi_{(i)}$ using rule R_2.

Theorem 3.1. For $\underline{\lambda} \in \Omega$ and $i < j$, $p_{\underline{\lambda}}(i) \geq p_{\underline{\lambda}}(j)$.

Proof:

$$p_{\underline{\lambda}}(i) = P_{\underline{\lambda}}(\text{select population } \pi_{(i)} | R_2)$$

$$= P_{\underline{\lambda}}(X_{(i)} \leq c_2(t) \min_{\ell \neq i} X_{(\ell)} + c_2(t) | \sum_{\ell=1}^{k} X_{(\ell)} = t)$$

$$= \sum_{\substack{x_1, \ldots, \hat{x}_i, \ldots, \hat{x}_j, \ldots, x_k \\ x_i \leq c_2(t) \min_{\ell \neq i} x_\ell + c_2(t)}} \sum_{x_i \leq \frac{c_2(t)(x_i+x_j+1)}{1+c_2(t)}} \binom{x_i+x_j}{x_i} \left(\frac{p_i}{p_i+p_j}\right)^{x_i} \left(\frac{p_j}{p_i+p_j}\right)^{x_j} t! \frac{(p_i+p_j)^{x_i+x_j}}{(x_i+x_j)!} \prod_{\substack{\ell=1 \\ \ell \neq i \\ \ell \neq j}}^{k} \frac{q_\ell^{x_\ell}}{x_\ell!} .$$

where $p_i = \dfrac{\lambda_{[i]}}{\sum_{\ell=1}^{k} \lambda_{[\ell]}}$, $q_r = \dfrac{\lambda_{[r]}}{\sum_{\substack{\ell=1 \\ \ell \neq i \\ \ell \neq j}}^{k} \lambda_{[\ell]}}$ and $\hat{x}_i$ denote that x_i is deleted. Note that when x_i and x_j are interchanged, the second part in the above summand remains unchanged, and Binomial distribution belongs to the stochastically increasing family. So the result follows.

3.2. The Probability of a Correct Selection for R_2

Lemma 3.1. For k = 2,

$$\inf_{\underline{\lambda} \in \Omega} P_{\underline{\lambda}}(CS|R_2) = \inf_{\underline{\lambda} \in \Omega_0} P_{\underline{\lambda}}(CS|R_2).$$

Proof. For $\underline{\lambda} \in \Omega$,

$$P_{\underline{\lambda}}(CS|R_2) = P_{\underline{\lambda}}(X_{(1)} \leq c_2(t)X_{(2)} + c_2(t)|X_{(1)}+X_{(2)}=t)$$

$$= \sum_{x=0}^{\left[\frac{c_2(t)(1+t)}{1+c_2(t)}\right]} \binom{t}{x}\left(\frac{\lambda_{[1]}}{\lambda_{[1]}+\lambda_{[2]}}\right)^x\left(\frac{\lambda_{[2]}}{\lambda_{[1]}+\lambda_{[2]}}\right)^{t-x}.$$

For fixed $\lambda_{[2]}$, $\frac{\lambda_{[1]}}{\lambda_{[1]}+\lambda_{[2]}}$ increases with $\lambda_{[1]}$ to $\frac{1}{2}$, this implies that

$$\inf_{\underline{\lambda} \in \Omega} P_{\underline{\lambda}}(CS|R_2) = \inf_{\lambda \in \Omega_0} P_{\underline{\lambda}}(CS|R_2).$$

Theorem 3.2. For a given P^*, $\frac{1}{k} < P^* < 1$, $k = 2$ and any $t \geq 0$, let $c_2(t)$ be the smallest value such that

$$P_{\Omega_0}\left(X_1 \leq \frac{c_2(t)(1+t)}{1+c_2(t)} \,\middle|\, X_1+X_2=t\right) \geq P^*.$$

Then $\inf_{\underline{\lambda} \in \Omega} P_{\underline{\lambda}}(CS|R_2) \geq P^*$.

The result follows immediately from Lemma 3.1.

For $k \geq 3$, we need the following definitions in order to discuss the least favorable configuration of $P_{\underline{\lambda}}(CS|R_2)$.

Definition 3.1. If $a_{[1]} \leq a_{[2]} \leq \ldots \leq a_{[m]}$, and $b_{[1]} \leq b_{[2]} \leq \ldots \leq b_{[m]}$ denote the ordered values of the components of $\underline{a}$ and $\underline{b}$, respectively, and such that

$\sum_{i=1}^{r} a_{[m-i+1]} \geq \sum_{i=1}^{r} b_{[m-i+1]}$, for $r = 1,2,\ldots,m-1$, and $\sum_{i=1}^{m} a_{[i]} = \sum_{i=1}^{m} b_{[i]}$,

then $\underline{a}$ is said to majorize $\underline{b}$, written $\underline{a} \succ \underline{b}$ or equivalently $\underline{b} \prec \underline{a}$.

Definition 3.2. If a function φ satisfies the property that $\varphi(\underline{x}) \leq \varphi(\underline{y})$ ($\varphi(\underline{x}) \geq \varphi(\underline{y})$) whenever $\underline{x} \succ \underline{y}$, then φ is called a Schur-concave (Schur-convex) function.

The following lemma is due to Rinott [7], and is stated without proof.

Lemma 3.2. Let $\underline{X} = (X_1,\ldots,X_k)$ have the multinomial distribution

$$P(\underline{X} = \underline{x}) = \binom{N}{x_1 \ldots x_k} \prod_{i=1}^{k} \theta_i^{x_i}$$

when $\underline{x} = (x_1,\ldots,x_k)$, $\sum_{i=1}^{k} x_i = N$ and $\sum_{i=1}^{k} \theta_i = 1$. Let $\phi(\underline{x})$ be a Schur function. Then $E_{\underline{\theta}}\phi(\underline{X})$ is a Schur-function.

Let $\Omega_2 = \{\underline{\lambda}=(\lambda_1,\ldots,\lambda_k)\colon 0 < \lambda_{[1]} = \ldots = \lambda_{[k-1]} < \lambda_{[k]}\}$.

Theorem 3.3.

$$\inf_{\underline{\lambda}\in\Omega} p_{\lambda}(CS|R_2) = \inf_{\underline{\lambda}\in\Omega_2} P_{\underline{\lambda}}(CS|R_2).$$

Proof. For $\underline{\lambda} \in \Omega$,

$$P_{\underline{\lambda}}(CS|R_2) = P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) \min_{2\leq j\leq k} X_{(j)} + c_2(t) \mid \sum_{i=1}^{k} X_i = t)$$

$$= \sum_{y_1=0}^{t} \binom{t}{y_1} p_1^{y_1} (1-p_1)^{t-y_1} \cdot \sum \binom{t-y_1}{y_2 \ldots y_k} \prod_{j=2}^{k} \left(\frac{p_j}{1-p_1}\right)^{y_j}$$

where $p_i = \lambda_{[i]}\left(\sum_{j=1}^{k} \lambda_{[j]}\right)^{-1}$, $i = 1,\ldots,k$ and the second summation is over the set of all $(k-1)$-tuples of nonnegative integers $(y_2,\ldots,y_k)$ such that $y_j \geq \frac{y_1 - c_2(t)}{c_2(t)}$, $j = 2,\ldots,k$ and $\sum_{j=2}^{k} y_j = t-y_1$. Let

$$\phi_{y_1}(y_2,\ldots,y_k) = \begin{cases} 1 \text{ if } y_j \geq \dfrac{y_1-c_2(t)}{c_2(t)}, & j=2,\ldots,k, \\ 0 \text{ otherwise.} \end{cases}$$

It is easy to see that $P_{\underline{\lambda}}(CS|R_2)$ can be written as $E_{Y_1}(E(\phi_{Y_1}(Y_2,\ldots,Y_k)|Y_1))$ where $(Y_1,\ldots,Y_k)$ is a multinomial random vector with parameters t and $(p_1,\ldots,p_k)$. Since for fixed y_1, $\phi_{y_1}(y_2,\ldots,y_k)$ is a Schur-concave function in $(y_2,\ldots,y_k)$, hence by Lemma 3.1, $P_{\underline{\lambda}}(CS|R_2)$ is Schur-concave in $(\frac{p_2}{1-p_1},\ldots,\frac{p_k}{1-p_1})$ when p_1 is kept fixed. This implies that $P_{\underline{\lambda}}(CS|R_2)$ is minimized when $p_1 = \ldots = p_{k-1}$, $p_k = 1-(k-1)p_1$, or when $\lambda_{[1]} = \ldots = \lambda_{[k-1]} < \lambda_{[k]}$. Thus the proof is completed.

Under the parameter space Ω_2, the joint distribution of $X_1,\ldots,X_k$ given $\sum_{i=1}^{k} X_i = t$, is a multinomial distribution with parameters t and $(p_1,\ldots,p_k)$ where $p_1 = \ldots = p_{k-1} = \frac{\lambda}{(k-1)\lambda+\lambda'} = p$, $p_k = \frac{\lambda'}{(k-1)\lambda+\lambda'} = q$, $p < q$.

<u>Theorem 3.4.</u>

$$\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_2) = \inf_{0<\lambda<\lambda'} \sum_{\substack{x_1 \leq c_2(t)\min_{j\neq 1} x_j + c_2(t) \\ x_i \geq 0,\ \Sigma x_i = t}} \binom{t}{x_1\ldots x_k}\left(\frac{1}{k-1+\frac{\lambda'}{\lambda}}\right)^t \left(\frac{\lambda'}{\lambda}\right)^{x_k}.$$

<u>Proof</u>. For $\underline{\lambda} \in \Omega_2$

$$P_{\underline{\lambda}}(CS|R_2) = P_{\underline{\lambda}}(X_{(1)} \leq c_2(t)\min_{j\neq 1} X_{(j)} + c_2(t) | \sum_{i=1}^{k} X_i = t)$$

$$= \sum_{\substack{x_1 \leq c_2(t)\min_{j\neq 1} x_j + c_2(t) \\ x_i \geq 0,\ \Sigma x_i = t}} \binom{t}{x_1\ldots x_k}\left(\frac{\lambda}{(k-1)\lambda+\lambda'}\right)^{\sum_1^{k-1} x_i}\left(\frac{\lambda'}{(k-1)\lambda+\lambda'}\right)^{x_k}.$$

The theorem follows from Theorem 3.3 after simplification.

<u>Theorem 3.5</u>. For $k \geq 3$, and for any P^*, let $P_2^* = 1 - \frac{1-P^*}{k-1}$, $0 \leq r \leq t$, let $c_2(r)$ be the smallest value such that

$$\sum_{i=0}^{\left[\frac{c_2(r)(1+r)}{1+c_2(r)}\right]} \binom{r}{i} \frac{1}{2^r} \geq P_2^*.$$

If $c_2(t) = \max\{c_2(r): 0 \leq r \leq t\}$, then $\inf_{\underline{\lambda} \in \Omega} P_{\underline{\lambda}}(CS|R_2) \geq P^*$.

<u>Proof</u>. For $\underline{\lambda} \in \Omega_1$,

$$P_{\underline{\lambda}}(CS|R_2) = P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) \min_{2 \leq j \leq k} X_{(j)} + c_2(t) | \sum_{i=1}^{k} X_{(i)} = t)$$

$$\geq 1 - \sum_{j=2}^{k} \{1 - P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) X_{(j)} + c_2(t) | \sum_{i=1}^{k} X_{(i)} = t)\}$$

$$= 2-k + \frac{1}{P_{\underline{\lambda}}(\sum_{i=1}^{k} X_{(i)} = t)} \sum_{j=2}^{k} \sum_{r=0}^{t} P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) X_{(j)} + c_2(t), X_{(1)} + X_{(j)} = r, \sum_{\substack{i \neq 1 \\ i \neq j}} X_{(i)} = t-r)$$

$$= 2-k + \frac{1}{P_{\underline{\lambda}}(\sum_{i=1}^{k} X_{(i)} = t)} \sum_{j=2}^{k} \sum_{r=0}^{t} P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) X_{(j)} + c_2(t), X_{(1)} + X_{(j)} = r) P_{\underline{\lambda}}(\sum_{\substack{i \neq 1 \\ i \neq r}} X_{(i)} = t-r)$$

$$= 2-k + \frac{1}{P_{\underline{\lambda}}(\sum_{i=1}^{k} X_{(i)} = t)} \sum_{j=2}^{k} \sum_{r=0}^{t} P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) X_{(j)} + c_2(t) | X_{(1)} + X_{(j)} = r) P_{\underline{\lambda}}(X_{(1)} + X_{(j)} = r) P_{\underline{\lambda}}(\sum_{\substack{i \neq 1 \\ i \neq j}} X_{(i)} = t-r)$$

$$= 2-k+ \frac{1}{P_{\underset{\sim}{\lambda}}(\Sigma X_{(i)}=t)} \sum_{j=2}^{k} \sum_{r=0}^{t} \sum_{i=0}^{\left[\frac{c_2(t)(r+1)}{1+c_2(t)}\right]} \binom{r}{i}\left(\frac{\lambda_{[1]}}{\lambda_{[1]}+\lambda_{[j]}}\right)^i\left(\frac{\lambda_{[j]}}{\lambda_{[1]}+\lambda_{[j]}}\right)^{r-i}\} .$$

$$P_{\underset{\sim}{\lambda}}(X_{(1)}+X_{(j)}=r)P_{\underset{\sim}{\lambda}}(\sum_{\substack{i\neq 1\\ i\neq j}} X_{(i)}=t-r)$$

$$\geq 2-k+ \frac{1}{P_{\underset{\sim}{\lambda}}(\sum_{i=1}^{k} X_{(i)}=t)} \sum_{j=2}^{k} \sum_{r=0}^{t} \{ \sum_{i=0}^{\left[\frac{c_2(t)(1+r)}{1+c_2(t)}\right]} \binom{r}{i} \frac{1}{2^r}\} P_{\underset{\sim}{\lambda}}(X_{(1)}+X_{(j)}=r)P_{\underset{\sim}{\lambda}}(\sum_{\substack{i\neq 1\\ i\neq j}} X_{(i)}=t-r)$$

$$\geq 2-k+(k-1)P_2^*$$

$$= P^*.$$

Thus we have the result.

Hence, for each k and P*, Theorem 3.5. guarantees the existence of $c_2(t)$ and gives a method to find $c_2(t)$ for given $\sum_{i=1}^{k} X_i=t$ such that $P_{\underset{\sim}{\lambda}}(CS|R_2) \geq P^*$ for any $\underset{\sim}{\lambda} \in \Omega$.

3.3 An Upper Bound on the Expected Subset Size for R_2

For any fixed values of k and P*, the expected size of the selected subset by using procedure R_2 is a function of the true configuration $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_k)$. Now consider the space of all slippage configurations of the type $\lambda_{[1]} = \delta\lambda$, $\delta < 1$, $\lambda_{[2]} =\ldots= \lambda_{[k]} = \lambda$, $\lambda > 0$. Let us denote this space by Ω_3.

Theorem 3.6.

$$\sup_{\underset{\sim}{\lambda}\in\Omega_3} E_{\underset{\sim}{\lambda}}(S|R_2) \leq k- \sum_{r=0}^{t} \sum_{s=0}^{\left[\frac{r-c_2(t)}{1+c_2(t)}\right]-1} \binom{r}{s}\{\delta^{r-s}+(k-1)\delta^s\}\binom{t}{r}\frac{(k-2)^{t-r}}{(k-1+\delta)^t} .$$

Proof. For any $\underline{\lambda} \in \Omega_3$,

$$E_{\underline{\lambda}}(S|R_2) = P_{\underline{\lambda}}(X_{(1)} \leq c_2(t) \min_{2\leq j\leq k} X_{(j)} + c_2(t) | \sum_{i=1}^{k} X_i = t)$$

$$+ (k-1)P_{\underline{\lambda}}(X_{(2)} \leq c_2(t) \min_{j\neq 2} X_{(j)} + c_2(t) | \sum_{i=1}^{k} X_i = t)$$

$$\leq k - P_{\underline{\lambda}}(X_{(1)} > c_2(t) X_{(2)} + c_2(t) | \sum_{i=1}^{k} X_i = t) - (k-1)P_{\underline{\lambda}}(X_{(2)} > c_2(t) X_{(1)} + c_2(t) | \sum_{i=1}^{k} X_i = t)$$

$$\leq k - \frac{1}{P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i = t)} \{ \sum_{r=0}^{t} \{ \sum_{s=0}^{[\frac{r-c_2(t)}{1+c_2(t)}]-1} \binom{r}{s} (\frac{1}{1+\delta})^s (\frac{\delta}{1+\delta})^{r-s}$$

$$+ (k-1) \sum_{s=0}^{[\frac{r-c_2(t)}{1+c_2(t)}]-1} \binom{r}{s} (\frac{\delta}{1+\delta})^s (\frac{1}{1+\delta})^{r-s} \} P_{\underline{\lambda}}(X_{(1)} + X_{(2)} = r) \cdot$$

$$P_{\underline{\lambda}}(\sum_{3}^{k} X_{(i)} = t-r)\}.$$

$$= k - \{ \sum_{r=0}^{t} \{ \sum_{s=0}^{[\frac{r-c_2(t)}{1+c_2(t)}]-1} \binom{r}{s} (\delta^{r-s} + (k-1)\delta^s)(1+\delta)^{-r} \} \frac{t!}{(\delta+(k-1))^t} \cdot$$

$$\frac{(1+\delta)^r}{r!} \frac{(k-2)^{t-r}}{(t-r)!} .$$

After simplifying, we have the result.

4. Other Selction Procedures

4.1. The Selection Procedure R_3

In this section we consider a selection procedure of the type suggested by Seal [8].

R_3: Select population π_i if and only if

$$X_i \le c_3 + \frac{c_3}{k-1} \sum_{j\neq i} X_j \tag{4.1}$$

where $c_3 \ge 1$ is the smallest constant to be chosen so as to satisfy the basic probability requirement (1.1).

By using an analogous argument as in the proof of Theorem 2.1, we have the following theorem.

Theorem 4.1. $\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_3) = \inf_{\underline{\lambda}\in\Omega_0} P_{\underline{\lambda}}(CS|R_3)$.

Moreover, it is easy to prove the following result.

Theorem 4.2. For any P*, any t, t > 0, let $c_3(t)$ be the smallest value such that

$$\sum_{i=0}^{\left[\frac{(k-1)c_3(t)+tc_3(t)}{k-1+c_3(t)}\right]} \binom{t}{i}\left(\frac{1}{k}\right)^i\left(\frac{k-1}{k}\right)^{t-i} \ge P^*.$$

If $c_3 = \sup\{c_3(t): t \ge 0\}$, then $\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_3) \ge P^*$.

Consider the special configuration $\lambda_{[1]} = \delta\lambda$, $\delta < 1$; $\lambda_{[2]} = \ldots = \lambda_{[k]} = \lambda$, $\lambda > \lambda_0 > 0$. Using the same notation as in Section 2, the space of all such slippage configuration is denoted by Ω_1. In the following theorem, we give an upper bound for the expected subset size when the rule R_3 is used.

Theorem 4.3. $\sup_{\underline{\lambda}\in\Omega} E_{\underline{\lambda}}(S|R_3) \le \sup_{r\ge 2} h(r) + (k - \sup_{r\ge 2} h(r))(1+(k-1+\delta)\lambda_0)e^{-(k-1+\delta)\lambda_0}$

where

$$h(r) = \sum_{i=0}^{\left[\frac{c_3(r+k-1)}{c_3+k-1}\right]} \binom{r}{i}\left\{\left(\frac{\delta}{k-1+\delta}\right)^i\left(1-\frac{\delta}{k-1+\delta}\right)^{r-i} + (k-1)\left(\frac{1}{k-1+\delta}\right)^i\left(1-\frac{1}{k-1+\delta}\right)^{r-i}\right\}.$$

Proof. For any $\underline{\lambda} \in \Omega_1$,

$$E_{\underline{\lambda}}(S|R_3) = P_{\underline{\lambda}}(X_{(1)} \leq c_3 + \frac{c_3}{k-1} \sum_{j=2}^{k} X_{(j)}) + (k-1) P_{\underline{\lambda}}(X_{(k)} \leq c_3 + \frac{c_3}{k-1} \sum_{i=1}^{k-1} X_{(i)})$$

$$= \sum_{r=0}^{\infty} \{P_{\underline{\lambda}}(X_{(1)} \leq c_3 + \frac{c_3}{k-1} \sum_{j=2}^{k} X_{(j)} | \sum_{i=1}^{k} X_i = r) + (k-1) P_{\underline{\lambda}}(X_{(k)} \leq c_3 + \frac{c_3}{k-1} \sum_{i=1}^{k-1} X_{(i)} | \sum_{i=1}^{k} X_i = r)\} P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i = r).$$

$$= \sum_{r=0}^{\infty} \{P_{\underline{\lambda}}(X_{(1)} \leq \frac{c_3(r+k-1)}{c_3+k-1} | \sum_{i=1}^{k} X_i = r) + (k-1) P_{\underline{\lambda}}(X_{(k)} \leq \frac{c_3(r+k-1)}{c_3+k-1} | \sum_{i=1}^{k} X_i = r)\} \cdot P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i = r)$$

$$= \sum_{r=0}^{\infty} \sum_{i=0}^{\left[\frac{c_3(r+k-1)}{c_3+k-1}\right]} \binom{r}{i} \{(\frac{\delta}{k-1+\delta})^i (1-\frac{\delta}{k-1+\delta})^{r-i} + (k-1)(\frac{1}{k-1+\delta})^i (1-\frac{1}{k-1+\delta})^{r-i}\} \cdot e^{-(k-1+\delta)\lambda} \frac{((k-1+\delta)\lambda)^r}{r!}$$

$$\leq k\{e^{-(k-1+\delta)\lambda}(1+(k-1+\delta)\lambda)\} + \sup_{r \geq 2} h(r) \{\sum_{i=2}^{\infty} e^{-(k-1+\delta)\lambda} \frac{((k-1+\delta)\lambda)^i}{i!}\}$$

$$\leq \sup_{r \geq 2} h(r) + (k - \sup_{r \geq 2} h(r)) e^{-(k-1+\delta)\lambda_0} (1+(k-1+\delta)\lambda_0).$$

The proof is completed.

4.2. A Conditional Selection Procedure R_4

We consider a conditional procedure as follows:

R_4: Select the population π_i if and only if

$$X_i \leq c_4(t) + \frac{c_4(t)}{k-1} \sum_{j \neq i} X_j \quad \text{given} \quad \sum_{i=1}^{k} X_i = t. \tag{4.2}$$

We know that the conditional distribution of $(X_1,\dots,X_k)$ given $\sum_{i=1}^{k} X_i = t$ is a multinomial distribution with parameters t and $\left(\frac{\lambda_1}{\sum_{i=1}^{k}\lambda_i},\dots,\frac{\lambda_k}{\sum_{i=1}^{k}\lambda_i}\right)$.

Theorem 4.4. $\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_4) = \inf_{\underline{\lambda}\in\Omega_0} P_{\underline{\lambda}}(CS|R_4)$.

Proof. For $\underline{\lambda} \in \Omega$,

$$P_{\underline{\lambda}}(CS|R_4) = P_{\underline{\lambda}}\left(X_{(1)} \le c_4(t) + \frac{c_4(t)}{k-1}\sum_{j=2}^{k} X_{(j)} \,\Big|\, \sum_{i=1}^{k} X_i = t\right) \tag{4.3}$$

$$= P_{\underline{\lambda}}\left(X_{(1)} \le \frac{c_4(t)(t+k-1)}{c_4(t)+k-1} \,\Big|\, \sum_{i=1}^{k} X_i = t\right)$$

$$= \sum_{i=0}^{\left[\frac{c_4(t)(t+k-1)}{c_4(t)+k-1}\right]} \binom{t}{i}\left(\frac{\lambda_{[1]}}{\sum_{j=1}^{k}\lambda_{[j]}}\right)^i \left(1-\frac{\lambda_{[1]}}{\sum_{j=1}^{k}\lambda_{[j]}}\right)^{t-i}.$$

The right hand member of (4.3) will be minimized when $\lambda_{[1]}=\dots=\lambda_{[k]}=\lambda$. In this case

$$\inf_{\underline{\lambda}\in\Omega} P_{\underline{\lambda}}(CS|R_4) = \inf_{\underline{\lambda}\in\Omega_0} P_{\underline{\lambda}}(CS|R_4) = \sum_{i=0}^{\left[\frac{c_4(t)(t+k-1)}{c_4(t)+k-1}\right]} \binom{t}{i}\left(\frac{1}{k}\right)^i\left(\frac{k-1}{k}\right)^{t-i}.$$

Note the infimum of the probability of a correct selection is independent of the common value λ and $c_4(t)$ is the smallest constant determined from the following inequality.

$$\sum_{i=0}^{\left[\frac{c_4(t)(t+k-1)}{c_4(t)+k-1}\right]} \binom{t}{i}(k-1)^{t-i} \ge k^t P^*.$$

Theorem 4.5. For any $\underline{\lambda} \in \Omega_1$,

$$E_{\underline{\lambda}}(S|R_4) = \sum_{i=0}^{[D(t)]} \binom{t}{i}\left\{\left(\frac{\delta}{k-1+\delta}\right)^i\left(1-\frac{\delta}{k-1+\delta}\right)^{t-i}+(k-1)\left(\frac{1}{k-1+\delta}\right)^i\left(1-\frac{1}{k-1+\delta}\right)^{t-i}\right\}$$

where $D(t) = \dfrac{c_4(t)(t+k-1)}{c_4(t)+k-1}$.

Proof. For $\underline{\lambda} \in \Omega_1$,

$$E_{\underline{\lambda}}(S|R_4)=P_{\underline{\lambda}}\left(X_{(1)}\le c_4(t)+\frac{c_4(t)}{k-1}\sum_{i=2}^{k}X_{(i)}\,\Big|\,\sum_{i=1}^{k}X_i=t\right)+(k-1)P_{\underline{\lambda}}\left(X_{(k)}\le c_4(t)+\frac{c_4(t)}{k-1}\sum_{i=1}^{k-1}X_{(i)}\,\Big|\,\sum_{i=1}^{k}X_i=t\right)$$

$$= P_{\underline{\lambda}}\left(X_{(1)}\le D(t)\,\Big|\,\sum_{i=1}^{k}X_i=t\right)+(k-1)P_{\underline{\lambda}}\left(X_{(k)}\le D(t)\,\Big|\,\sum_{i=1}^{k}X_i=t\right).$$

The theorem follows easily.

5. Applications to a Test of Homogeneity for $\lambda_1 = \ldots = \lambda_k$.

In some practical situations one wishes to know whether λ_i are significantly different or not. This is the problem of the test of homogeneity of the Poisson populations. In order to test the homogeneity of k populations, i.e. to test H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_k = \lambda_0$ against the alternative H_A: not A, we propose the following rule ϕ_1 and $\phi_2(T)$.

(1) The procedure ϕ_1: H_0 is accepted if, and only if $X_{max}-cX_{min} \le c$ where c is some constant depending on k, λ_0 and the level of significance α.

(2) The procedure $\phi_2(T)$: H_0 is accepted if, and only if

$$X_{max}-c(t)X_{min} \le c(t), \text{ given } T = \sum_{i=1}^{k} X_i=t.$$

For the procedure ϕ_1, if we choose $c = \sup\{c(t):\ t \ge 0\}$, where for any t, $t \ge 0$ c(t) is the smallest constant such that

$$A(k,t,c(t)) \geq 1 - \frac{\alpha}{k},$$

then under H_0,

$$P_{\underline{\lambda}}(X_{max} - cX_{min} \leq c)$$

$$= 1 - P_{\underline{\lambda}}(\max_{1 \leq i \leq k} X_i > c \min_{1 \leq j \leq k} X_j + c)$$

$$\geq 1 - \sum_{i=1}^{k} P_{\underline{\lambda}}(X_i > c \min X_j + c)$$

$$= 1 - k + \sum_{i=1}^{k} P_{\underline{\lambda}}(X_i \leq c \min X_j + c)$$

$$= 1 - k + k \sum_{t=0}^{\infty} P_{\underline{\lambda}}(X_1 \leq c \min_{2 \leq j \leq k} X_j + c \mid \sum_{i=1}^{k} X_i = t) P_{\underline{\lambda}}(\sum_{i=1}^{k} X_i = t)$$

$$\geq 1 - k + k(1 - \frac{\alpha}{k})$$

$$= 1 - \alpha.$$

Hence $P_{H_0}(\text{Reject } H) \leq \alpha$.

Similarly, the probability of the error of the first kind for $\phi_2(T)$ is then given by

$$P(\max_{1 \leq j \leq k} X_j - c(t) \min_{1 \leq j \leq k} X_j > c(t) \mid \sum_{i=1}^{k} X_i = t)$$

$$= P_{\underline{\lambda}}(X_i - c(t) \min_{1 \leq j \leq k} X_j > c(t) \text{ for some } i \mid \sum_{i=1}^{k} X_i = t)$$

$$\leq \sum_{i=1}^{k} P_{\underline{\lambda}}(X_i > c(t) \min_{1 \leq j \leq k} X_j + c(t) \mid \sum_{i=1}^{k} X_i = t)$$

$$= k\{1 - P_{\underline{\lambda}}(X_1 \leq c(t) \min_{2 \leq j \leq 1} X_j + c(t) \mid \sum_{i=1}^{k} X_i = t)\}$$

$$= k(1 - A(k,t,c(t)))$$

$$\leq k(1 - (1 - \frac{\alpha}{k}))$$

$$= \alpha.$$

7. Explanations of the Tables

(1) Tables I and II list the infimum of the probability of a correct selection (approximate value) for the rules R_1 and R_3. It should be pointed out that the probability of a correct selection for these rules is decreasing when λ is small and then it is increasing again with λ. Hence, the approximate infimum can be determined numerically by computing the probability as a function of λ, for fixed values of c. For given k and P*, the selection constants (approximately) can be found from these tables. For example, for P* = .8504 and k = 4, the approximate value of c associated with R_1 is 2.4.

(2) In tables IIIA, IIIB, IIIC and IIID, the first entry denotes the probability of selecting the best population, the second entry denotes the probability of selecting a non-best population and the third entry is the expected proportion, all under the slippage configuration $\lambda_{[1]} = \delta\lambda$, $\delta < 1$; $\lambda_{[2]} = \ldots = \lambda_{[k]} = \lambda$, when the rule R_1 is used. The three entries in Table IVA, IVB, IVC, IVD define the same quantities for the rule R_3. For example, from Table IIIC, we find that for the rule R_1 if $\lambda = 2.00$ and c = 1.50 (k = 5 and $\delta = 0.3$), the probability of a correct selection is .9447, the probability of selecting a non-best population is .5399 and the expected proportion of populations in the selected subset is .6208.

8. Some Remarks on the Comparison of R_1 and R_3

We define a rule R to be better than another rule R' if the expected proportion for R is smaller than the expected proportion for R'. We compare the performance of the rules R_1 and R_3 in this aspect. For example, when k = 5, P* = 0.92, we obtain the approximate values of selection constants for R_1 and R_3 as $c_1 = 3.0$, $c_3 = 1.6$ from Table I and Table II respectively. For this

constants Tables III, IV show that if δ is kept fixed, R_3 seems to be better than R_1 when λ is small, while R_1 performs better than R_3 for large values of λ.

Table I

Table of inf $P(CS|R_1)$ (Approximate) using the Rule R_1

k \ c	1.6	1.8	2.0	2.4	2.8	3.0	3.5	4.0	4.5	5.0
2	0.8577	0.8762	0.9353	0.9391	0.9517	0.9771	0.9792	0.9902	0.9906	0.9956
3	0.7627	0.7895	0.8845	0.8904	0.9118	0.9566	0.9604	0.9811	0.9817	0.9913
4	0.6936	0.7246	0.8431	0.8504	0.8784	0.9380	0.9433	0.9724	0.9733	0.9872
5	0.6394	0.6740	0.8076	0.8151	0.8484	0.9209	0.9277	0.9643	0.9654	0.9832
6	0.5963	0.6313	0.7769	0.7845	0.8212	0.9053	0.9135	0.9566	0.9578	0.9793
8	0.5322	0.5644	0.7263	0.7341	0.7750	0.8774	0.8881	0.9425	0.9439	0.9720
10	0.4807	0.5144	0.6858	0.6943	0.7374	0.8532	0.8641	0.9289	0.9314	0.9651

For given k and c, this table represents the minimum value (approximately) of

$$P_\lambda[X_k \leq c \min_{1 \leq j \leq k-1} X_j + c] = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} \left\{ \sum_{j=<\frac{i}{c}-1>}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \right\}^{k-1}$$

where $X_1, \ldots, X_k$ are i.i.d. Poisson variables with parameter λ.

Table II

Table of inf $P(CS|R_3)$ (Approximate) using the Rule R_3

k \ c	1.6	1.8	2.0	2.4	2.8	3.0	3.5	4.0	4.5	5.0
2	0.8577	0.8762	0.9353	0.9391	0.9517	0.9771	0.9792	0.9902	0.9906	0.9956
3	0.8996	0.9407	0.9575	0.9772	0.9887	0.9950	0.9965	0.9989	0.9990	0.9996
4	0.9201	0.9452	0.9730	0.9826	0.9937	0.9953	0.9985	0.9995	0.9997	0.9999
5	0.9260	0.9573	0.9733	0.9889	0.9955	0.9979	0.9993	0.9995	0.9998	0.9999
6	0.9389	0.9611	0.9796	0.9911	0.9964	0.9982	0.9993	0.9998	0.9999	0.9999
8	0.9453	0.9676	0.9828	0.9938	0.9973	0.9987	0.9995	0.9999	0.9999	0.9999
10	0.9465	0.9678	0.9845	0.9940	0.9981	0.9987	0.9997	0.9999	0.9999	0.9999

For given k and c_3, this table represents the minimum value (approximately) of

$$P_\lambda[X_k \leq \frac{c_3}{k-1}\sum_{j=1}^{k-1} X_j + c_3] = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!}\{\sum_{j=<(k-1)(\frac{i}{c_3}-1)>}^{\infty} e^{-(k-1)\lambda}\frac{((k-1)\lambda)^j}{j!}\} = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!}$$

$$\{\int_0^{(k-1)\lambda} \frac{1}{\Gamma(<(k-1)(\frac{i}{c_3}-1)>)} y^{<(k-1)(\frac{i}{c_3}-1)>-1} e^{-y} dy\}$$

where $X_1,\ldots,X_k$ are i.i.d. Poisson variables with parameter λ.

Table IIIA

Using the rule R_1 and under the configuration $(\delta\lambda, \lambda, \ldots, \lambda)$, this tables gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 3,\ \delta = 0.3$

λ \ c_1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	0.9777	0.9978	0.9978	0.9998	0.9998	0.9999	0.9999	0.9999
	0.7761	0.9322	0.9327	0.9841	0.9841	0.9969	0.9969	0.9995
	0.8433	0.9541	0.9544	0.9893	0.9893	0.9979	0.9979	0.9996
2.0	0.9678	0.9940	0.9941	0.9991	0.9991	0.9999	0.9999	0.9999
	0.5889	0.7857	0.7974	0.9114	0.9125	0.9678	0.9678	0.9898
	0.7152	0.8551	0.8630	0.9406	0.9413	0.9785	0.9785	0.9932
3.0	0.9736	0.9932	0.9938	0.9986	0.9986	0.9997	0.9997	0.9999
	0.4880	0.6729	0.7146	0.8332	0.8415	0.9179	0.9190	0.9632
	0.6499	0.7797	0.8077	0.8883	0.8939	0.9452	0.9459	0.9755
4.0	0.9811	0.9944	0.9954	0.9987	0.9987	0.9997	0.9997	0.9999
	0.4111	0.5945	0.6680	0.7783	0.8020	0.8752	0.8803	0.9314
	0.6011	0.7278	0.7771	0.8518	0.8676	0.9167	0.9201	0.9542
5.0	0.9866	0.9960	0.9971	0.9990	0.9991	0.9997	0.9997	0.9999
	0.3481	0.5360	0.6307	0.7411	0.7822	0.8480	0.8609	0.9075
	0.5609	0.6893	0.7528	0.8271	0.8545	0.8986	0.9072	0.9383
6.0	0.9904	0.9973	0.9983	0.9993	0.9994	0.9998	0.9998	0.9999
	0.2980	0.4892	0.5961	0.7134	0.7679	0.8313	0.8537	0.8940
	0.5288	0.6586	0.7302	0.8087	0.8451	0.8874	0.9024	0.9293

Table IIIB

Using the rule R_1 and under the configuration $(\delta\lambda,\lambda,\ldots,\lambda)$, this tables gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 3,\ \delta = 0.5$

λ \ c_1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	0.9452	0.9913	0.9913	0.9989	0.9989	0.9998	0.9998	0.9999
	0.7871	0.9388	0.9395	0.9857	0.9857	0.9972	0.9972	0.9995
	0.8465	0.9563	0.9568	0.9901	0.9901	0.9981	0.9981	0.9996
2.0	0.9249	0.9782	0.9794	0.9951	0.9952	0.9990	0.9990	0.9998
	0.6679	0.8322	0.8467	0.9334	0.9347	0.9760	0.9761	0.9924
	0.7536	0.8809	0.8910	0.9540	0.9549	0.9837	0.9837	0.9949
3.0	0.9339	0.9762	0.9802	0.9933	0.9935	0.9981	0.9981	0.9995
	0.6066	0.7695	0.8138	0.8945	0.9029	0.9501	0.9512	0.9779
	0.7157	0.8384	0.8692	0.9274	0.9331	0.9661	0.9668	0.9851
4.0	0.9443	0.9798	0.9857	0.9941	0.9946	0.9980	0.9980	0.9993
	0.5570	0.7338	0.8028	0.8777	0.8977	0.9372	0.9413	0.9666
	0.6861	0.8158	0.8638	0.9165	0.9300	0.9574	0.9602	0.9775
5.0	0.9534	0.9842	0.9904	0.9957	0.9964	0.9984	0.9985	0.9994
	0.5187	0.7108	0.7940	0.8714	0.9007	0.9345	0.9431	0.9629
	0.6636	0.8019	0.8594	0.9128	0.9326	0.9558	0.9616	0.9751
6.0	0.9617	0.9881	0.9936	0.9971	0.9978	0.9989	0.9990	0.9995
	0.4897	0.6943	0.7871	0.8695	0.9037	0.9359	0.9483	0.9641
	0.6470	0.7923	0.8559	0.9120	0.9351	0.9569	0.9652	0.9759

Table IIIC

Using the rule R_1 and under the configuration $(\delta\lambda,\lambda,\ldots,\lambda)$, this table gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 5,\ \delta = 0.3$

λ \ c_1	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	0.9689	0.9969	0.9969	0.9997	0.9997	0.9999	0.9999	0.9999
	0.7518	0.9247	0.9249	0.9822	0.9822	0.9965	0.9965	0.9994
	0.7952	0.9391	0.9393	0.9857	0.9857	0.9972	0.9972	0.9995
2.0	0.9447	0.9896	0.9897	0.9985	0.9985	0.9998	0.9998	0.9999
	0.5399	0.7568	0.7665	0.8975	0.8985	0.9626	0.9627	0.9882
	0.6208	0.8034	0.8112	0.9177	0.9185	0.9700	0.9701	0.9906
3.0	0.9518	0.9874	0.9882	0.9975	0.9975	0.9995	0.9995	0.9999
	0.4487	0.6408	0.6827	0.8135	0.8221	0.9078	0.9088	0.9587
	0.5493	0.7101	0.7438	0.8503	0.8572	0.9261	0.9270	0.9669
4.0	0.9648	0.9894	0.9910	0.9975	0.9975	0.9994	0.9994	0.9998
	0.3812	0.5671	0.6437	0.7612	0.7865	0.8654	0.8708	0.9260
	0.4980	0.6516	0.7132	0.8084	0.8287	0.8922	0.8966	0.9408
5.0	0.9748	0.9923	0.9943	0.9981	0.9982	0.9995	0.9998	0.9998
	0.3242	0.5139	0.6122	0.7278	0.7715	0.8406	0.8643	0.9032
	0.4542	0.6096	0.6886	0.7819	0.8168	0.8724	0.9321	0.9225
6.0	0.9818	0.9948	0.9966	0.9987	0.9988	0.9996	0.9996	0.9998
	0.2794	0.4718	0.5817	0.7034	0.7603	0.8261	0.8496	0.8912
	0.4199	0.5764	0.6647	0.7624	0.8080	0.8608	0.8796	0.9129

Table IIID

Using the rule R_1 and under the configuration $(\delta\lambda,\lambda,...,\lambda)$, this table gives in order the tripl (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 5,\ \delta = 0.5$

λ \ c	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	0.9239	0.9878	0.9878	0.9985	0.9985	0.9998	0.9998	0.9999
	0.7601	0.9273	0.9276	0.9829	0.9828	0.9967	0.9967	0.9994
	0.7929	0.9394	0.9396	0.9860	0.9860	0.9973	0.9973	0.9995
2.0	0.8748	0.9629	0.9643	0.9915	0.9916	0.9983	0.9983	0.9997
	0.5954	0.7899	0.8029	0.9139	0.9151	0.9687	0.9688	0.9902
	0.6513	0.8245	0.8351	0.9294	0.9304	0.9747	0.9747	0.9921
3.0	0.8861	0.9569	0.9628	0.9875	0.9878	0.9965	0.9965	0.9991
	0.5415	0.7199	0.7684	0.8674	0.8770	0.9367	0.9379	0.9718
	0.6104	0.7673	0.8073	0.8914	0.8992	0.9486	0.9496	0.9773
4.0	0.9030	0.9623	0.9724	0.9886	0.9895	0.9961	0.9962	0.9988
	0.4971	0.6869	0.7654	0.8531	0.8769	0.9244	0.9294	0.9598
	0.5783	0.7420	0.8068	0.8802	0.8995	0.9387	0.9427	0.9676
5.0	0.9171	0.9701	0.9814	0.9916	0.9929	0.9969	0.9970	0.9988
	0.4634	0.6687	0.7622	0.8512	0.8856	0.9247	0.9349	0.9576
	0.5541	0.7289	0.8060	0.8793	0.9071	0.9391	0.9473	0.9659
6.0	0.9306	0.9772	0.9875	0.9944	0.9957	0.9979	0.9981	0.9991
	0.4402	0.6573	0.7599	0.8534	0.8923	0.9288	0.9429	0.9605
	0.5383	0.8213	0.8055	0.8816	0.9130	0.9426	0.9539	0.9682

Table IVA

Using the rule R_3 and under the configuration $(\delta\lambda,\lambda,\ldots,\lambda)$, this tables gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 3,\ \delta = 0.3$

λ \ c_3	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	.9939	.9994	.9995	.9999	.9999	.9999	.9999	.9999
	.8931	.9701	.9766	.9945	.9947	.9989	.9990	.9998
	.9267	.9799	.9842	.9963	.9965	.9993	.9993	.9998
2.0	.9948	.9991	.9995	.9999	.9999	.9999	.9999	.9999
	.8176	.9227	.9605	.9848	.9881	.9958	.9960	.9987
	.8766	.9482	.9735	.9898	.9921	.9972	.9973	.9991
3.0	.9963	.9993	.9997	.9999	.9999	.9999	.9999	.9999
	.7590	.8828	.9511	.9784	.9875	.9945	.9957	.9981
	.8381	.9216	.9673	.9856	.9916	.9963	.9971	.9987
4.0	.9975	.9995	.9999	.9999	.9999	.9999	.9999	.9999
	.7235	.8599	.9419	.9739	.9865	.9937	.9966	.9983
	.8149	.9064	.9612	.9826	.9910	.9958	.9977	.9988
5.0	.9985	.9997	.9999	.9999	.9999	.9999	.9999	.9999
	.7006	.8503	.9365	.9724	.9857	.9931	.9970	.9985
	.7999	.9001	.9577	.9816	.9905	.9954	.9980	.9990
6.0	.9991	.9998	.9999	.9999	.9999	.9999	.9999	.9999
	.6818	.8471	.9363	.9734	.9864	.9933	.9972	.9986
	.7876	.8980	.9575	.9822	.9909	.9955	.9981	.9990

Table IVB

Using the rule R_3 and under the configuration $(\delta\lambda,\lambda,\ldots,\lambda)$, this table gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 3,\ \delta = 0.5$

λ \ c_3	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	.9830	.9975	.9980	.9997	.9997	.9999	.9999	.9999
	.9068	.9743	.9806	.9955	.9957	.9991	.9991	.9998
	.9322	.9820	.9864	.9969	.9970	.9994	.9994	.9999
2.0	.9819	.9958	.9981	.9995	.9996	.9999	.9999	.9999
	.8472	.9377	.9711	.9891	.9919	.9971	.9973	.9991
	.8921	.9571	.9801	.9926	.9944	.9980	.9981	.9994
3.0	.9837	.9957	.9989	.9997	.9998	.9999	.9999	.9999
	.8059	.9111	.9663	.9858	.9923	.9967	.9976	.9989
	.8652	.9393	.9771	.9904	.9948	.9978	.9983	.9993
4.0	.9873	.9966	.9993	.9998	.9999	.9999	.9999	.9999
	.7860	.9005	.9622	.9942	.9923	.9965	.9983	.9991
	.8531	.9325	.9746	.9894	.9948	.9977	.9988	.9994
5.0	.9908	.9978	.9995	.9998	.9999	.9999	.9999	.9999
	.7743	.8996	.9617	.9849	.9926	.9966	.9986	.9993
	.8464	.9323	.9743	.9899	.9950	.9977	.9990	.9995
6.0	.9934	.9986	.9997	.9999	.9999	.9999	.9999	.9999
	.7655	.9023	.9645	.9867	.9937	.9970	.9988	.9994
	.8414	.9344	.9762	.9911	.9958	.9980	.9992	.9996

Table IVC

Using the rule R_3 and under the configuration $(\delta\lambda, \lambda, \ldots, \lambda)$, this table gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected populations $([(a)+(k-1)(b)]/k)$.

$k = 5,\ \delta = 0.3$

λ \ c_3	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	.9977	.9990	.9998	.9999	.9999	.9999	.9999	.9999
	.9394	.9655	.9899	.9979	.9990	.9996	.9998	.9999
	.9510	.9722	.9919	.9983	.9992	.9997	.9998	.9999
2.0	.9979	.9994	.9999	.9999	.9999	.9999	.9999	.9999
	.8871	.9526	.9849	.9953	.9985	.9989	.9996	.9999
	.9092	.9620	.9879	.9962	.9988	.9991	.9997	.9999
3.0	.9987	.9998	.9999	.9999	.9999	.9999	.9999	.9999
	.8597	.9499	.9845	.9953	.9983	.9991	.9997	.9999
	.8875	.9599	.9876	.9962	.9986	.9993	.9998	.9999
4.0	.9993	.9999	.9999	.9999	.9999	.9999	.9999	1.0000
	.8508	.9519	.9865	.9963	.9987	.9994	.9998	.9999
	.8805	.9615	.9892	.9970	.9989	.9995	.9999	.9999
5.0	.9997	.9999	.9999	.9999	.9999	.9999	1.0000	1.0000
	.8462	.9555	.9892	.9973	.9992	.9997	.9999	.9999
	.8769	.9644	.9914	.9979	.9993	.9997	.9999	.9999
6.0	.9998	.9999	.9999	.9999	1.0000	1.0000	1.0000	1.0000
	.8437	.9596	.9914	.9981	.9995	.9998	.9999	.9999
	.8750	.9677	.9931	.9985	.9996	.9998	.9999	.9999

Table IVD

Using the rule R_3 and under the configuration $(\delta\lambda,\lambda,\ldots,\lambda)$, this tables gives in order the triple (a) the probability of selecting a best population, (b) the probability of selecting any non-best population and (c) the expected proportion of the selected population ([(a)+(k-1)(b)]/k).

k = 5, δ = 0.5

λ \ c_3	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1.0	.9915	.9961	.9993	.9999	.9999	.9999	.9999	.9999
	.9435	.9686	.9910	.9982	.9992	.9996	.9998	.9999
	.9531	.9741	.9927	.9985	.9993	.9997	.9998	.9999
2.0	.9898	.9972	.9995	.9999	.9999	.9999	.9999	.9999
	.8985	.9596	.9876	.9962	.9988	.9991	.9997	.9999
	.9168	.9672	.9900	.9970	.9990	.9993	.9998	.9999
3.0	.9917	.9985	.9997	.9999	.9999	.9999	.9999	.9999
	.8781	.9592	.9881	.9965	.9988	.9994	.9998	.9999
	.9009	.9671	.9904	.9972	.9990	.9995	.9995	.9999
4.0	.9945	.9992	.9999	.9999	.9999	.9999	.9999	.9999
	.8740	.9625	.9903	.9975	.9991	.9996	.9999	.9999
	.8981	.9698	.9922	.9980	.9993	.9997	.9999	.9999
5.0	.9964	.9996	.9999	.9999	.9999	.9999	.9999	.9999
	.8726	.9666	.9927	.9983	.9995	.9998	.9999	.9999
	.8974	.9732	.9941	.9986	.9996	.9998	.9999	.9999
6.0	.9976	.9998	.9999	.9999	.9999	.9999	1.0000	1.0000
	.8733	.9708	.9944	.9989	.9997	.9999	.9999	.9999
	.8982	.9766	.9955	.9991	.9998	.9999	.9999	.9999

References

[1] Bortkiewicz, L. von (1898). Das Gesetz der Kleinen Zahlen, Leipzig: Teubner.

[2] Chapman, D. G. (1952). On test and estimates for the ratio of Poisson means. Ann. Inst. Statist. Math., 4, 45-49.

[3] Goel, P. K. (1972). A note on the non-existence of subset selection procedure for Poisson populations. Mimeo. Ser. No. 303, Dept. of Statist., Purdue University, West Lafayette, Indiana 47907.

[4] Gupta, S. S. and Huang, D. Y. (1975). On subset selection procedures for Poisson populations and some applications to the multinomial selection problems. Applied Statistics (ed. Gupta, R. P.), North-Holland Publishing Co., Amsterdam, 97-109.

[5] Lehmann, E. L. (1959). Testing Statistical Hypothesis, John Wiley, New York.

[6] Leong, Y. K. and Wong, W. Y. (1977). On Poisson selection problems. Research Report 23/77, Dept. of Math., University of Malaya, K. Lumpur, Malaysia.

[7] Rinott, Y. (1973). Multivariate majorization and rearrangement inequalities with some application to probability and statistics, Israel J. Math., 15, 60-77.

[8] Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist., 26, 387-398.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Mimeograph Series #78-3	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On Subset Selection Procedures for Poisson Populations	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: Mimeo. Series #78-3
7. AUTHOR(s): Shanti S. Gupta, Yoon-Kwai Leong and Wing-Yue Wong	8. CONTRACT OR GRANT NUMBER(s): ONR N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, W. Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: May 1978
	13. NUMBER OF PAGES: 31
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Poisson Distributions, Ranking and Selection, Conditional Rules, Correct Selection, Expected Proportion, Test of Homogeneity.

sub i   ≥ or =   sub i

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper deals with the problem of selecting a subset containing the smallest parameter of $k(\geq 2)$ Poisson populations. The population parameters λ_i, $i=1,2,\ldots,k$ are assumed unknown and there is no a priori information about the correct pairing of the ordered and unordered λ_i's. Both unconditional and conditional selection rules are investigated. Tables are provided for approximate values of the constants necessary to carry out the procedures. Some other

next page

DD FORM 1473, 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE   S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

numerical computations have also been provided which shed light on the performance of the selection rule in terms of the probability of selecting a non-best population, the probability of a correct selection and the expected proportion in the selected subset. It should be pointed out that the problem treated here is not solvable by analogous methods for the problem of the maximum which was studied earlier by Gupta and Huang (1975).